

Topological reprogramming transforms an integral membrane oligosaccharyltransferase into a water-soluble glycosylation catalyst

Yong Hyun Kwon¹, Ljubica Mihaljević^{2,3}, Keehun Kim¹, David E. Kim^{2,3}, Thomas C. Donahue¹, Erik J. Bidstrup¹, Chandra K. Bandi¹, Belen Sotomayor¹, Sophia W. Hulbert⁴, Kathryn A. Myers⁵, Anru Tian⁵, Mariah Culpepper², Dario Mizrachi⁶, Thapakorn Jaroentomeechai^{7,8}, Henrik Clausen⁸, Michael C. Jewett⁵, David Baker^{2,3} and Matthew P. DeLisa^{1,4,9*}

¹Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, 120 Olin Hall, Ithaca, NY 14853, USA

²Department of Biochemistry, Institute for Protein Design, University of Washington, Seattle, WA 98195, USA

³Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA

⁴Biochemistry, Molecular and Cell Biology (BMCB), Cornell University, Ithaca, NY 14853, USA

⁵Department of Bioengineering, Stanford University, Stanford, CA 94305 USA

⁶Department of Physiology & Developmental Biology, Brigham Young University, Provo, UT 84602, USA

⁷Department of Biotechnology, Faculty of Science, Mahidol University, Bangkok 10400, Thailand

⁸Department of Cellular and Molecular Medicine, Faculty of Health Sciences, Copenhagen Center for Glycomics, University of Copenhagen, Copenhagen, Denmark

⁹Cornell Institute of Biotechnology, Cornell University, Ithaca, NY 14853, USA

*Address correspondence to: Matthew P. DeLisa, Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853. Tel: 607-254-8560; Email: md255@cornell.edu

Abstract

Glycosyltransferases (GTs) catalyze the formation of new glycosidic bonds and thus are vital for synthesizing nature's vast repertoire of glycans and glycoconjugates and for engineering glycan-related medicines and materials. However, obtaining detailed structural and functional insights for the >750,000 known GTs is limited by difficulties associated with their efficient recombinant expression. Members of the GT-C fold, in particular, pose the most significant expression challenges due to the integration and folding requirements of their multiple membrane-spanning regions. Here, we address this challenge by engineering water-soluble variants of an archetypal GT-C fold enzyme, namely the oligosaccharyltransferase PglB from *Campylobacter jejuni* (CjPglB), which possesses 13 hydrophobic transmembrane helices. To render CjPglB water-soluble, we leveraged two advanced protein engineering methods: one that is universal called SIMPLEX (solubilization of IMPs with high levels of expression) and the other that is custom tailored called WRAPs (water-soluble RFdiffused amphipathic proteins). Each approach was able to transform CjPglB into a water-soluble enzyme that could be readily expressed in the cytoplasm of *Escherichia coli* cells at yields in the 3–6 mg/L range. Importantly, solubilization was achieved without the need for detergents and with retention of catalytic function. Collectively, our findings demonstrate that both SIMPLEX and WRAPs are promising platforms for advancing the molecular characterization of even the most structurally complex GTs, while also enabling broader GT-mediated glycosylation capabilities within synthetic glycobiology applications.

Introduction

Glycosyltransferases (GTs) are ubiquitous enzymes that catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor substrates, resulting in the formation of glycosidic bonds¹. The diverse and extensive collection of GTs across all domains of life includes more than 750,000 known sequences^{2,3}, which encompass three main structural superfamilies—GT-A, GT-B, and GT-C—that are distinguished by unique folds and specific modes of substrate interaction^{1,4,5}. These enzymes use different substrates including nucleotide sugars (Leloir enzymes) and lipid-linked sugars (non-Leloir enzymes), and act on a range of acceptors including sugars, proteins, lipids, and small molecules. This broad enzymatic

capability represents a vital toolkit for glycosylation processes in nature¹⁻⁵ and biotechnology⁶⁻¹³.

Despite the vast repertoire, relatively few GTs have been characterized at the molecular level through structural and kinetic studies, leaving much of their molecular diversity poorly understood and restricting the full scope of their biotechnological potential. This gap has been ascribed to multiple challenges; however, one of the most significant, given its crucial role early in the characterization workflow, is the recombinant expression of GTs as functional proteins^{4, 5, 14-16}. One reason recombinant expression is so difficult stems from the fact that many GTs are membrane-associated, either via peripheral or single-pass anchoring in the case of GT-A and GT-B folds or multi-transmembrane helices in the case of GT-C folds. Consequently, many GTs are intrinsically hydrophobic and depend on membrane environments, leading to misfolding and aggregation following recombinant expression and poor *in vitro* solubility and stability following purification. While GT expression in mammalian, insect, and yeast cells has met some success, including two larger-scale expression campaigns^{15, 17}, there remains a need for a broadly effective expression platform that can be readily generalized for rapid, high-yield production of functional GTs from all three structural superfamilies.

To this end, we recently described a strategy for efficient production of structurally diverse GTs that leverages the biosynthetic capacity and versatility of *Escherichia coli* bacteria, the workhorse of modern molecular biology¹⁸. Our approach centered around SIMPLEX (solubilization of IMPs with high levels of expression), a universal protein engineering method for topologically converting peripheral and integral membrane proteins (IMPs) into water-soluble variants¹⁹. To adapt GTs for SIMPLEX, chimeras were created by genetically fusing candidate glycoenzymes with a decoy protein at their N-termini that prevented membrane insertion and an amphipathic protein at their C-termini that effectively shielded hydrophobic surfaces from the aqueous environment¹⁶. Our approach promoted soluble expression of nearly 100 GTs, including many of human origin, directly in the *E. coli* cytoplasm. Importantly, water-soluble GTs were all produced at high yields (5–10 mg/L) without the need for added detergents and with intact catalytic function.

In our previous work, we focused exclusively on enzymes from the GT-A and GT-B superfamilies, which use nucleotide sugars as donor substrates and only contain

between 0 and 1 transmembrane domains (TMDs) on average ¹⁶. In stark contrast, GT-C fold enzymes use lipid phosphate-activated donor sugar substrates and have between 8 and 13 TMDs with an active site located within a long-loop region ^{1, 20}, making them more topologically complex than other GT folds. Because of their complicated fold architectures, detailed experimental information about the structure and function of integral membrane GT-C fold enzymes is still sparse ²¹.

Therefore, in this work, we investigated whether the SIMPLEx-mediated solubilization approach could be used to generate water-soluble variants of GT-C superfamily members. As proof-of-concept, we focused on the PglB enzyme from *Campylobacter jejuni* (CjPglB), a bacterial single-subunit oligosaccharyltransferase (OST) that serves as an archetype of the GT-C fold, consisting of 13 TMDs and a conserved C-terminal globular domain that harbors the active site and the hallmark WWDYG motif critical for catalytic function ²². CjPglB and its structural homologues (e.g., archaeal AglB, eukaryotic STT3 catalytic subunit) mediate the key step of asparagine-linked (N-linked) glycosylation by transferring preassembled glycans from polyisoprenol-linked donors onto specific asparagine residues in acceptor proteins ^{23, 24}. Following several modifications to the original SIMPLEx design, we identified an optimized architecture that conferred water solubility to CjPglB in the absence of detergents, with final yields exceeding 5 mg/L. Importantly, the catalytic activity of SIMPLEx-solubilized CjPglB rivaled that of detergent-solubilized wild-type (wt) CjPglB, with both enzymes exhibiting ~100% glycosylation efficiency using a model acceptor protein substrate and exhibiting comparable Michaelis-Menten kinetics using a peptide substrate. In parallel, we also investigated an alternative deep learning-based design approach called WRAPs (water-soluble RFdiffused amphipathic proteins) ²⁵ for custom solubilization of CjPglB. Similar to SIMPLEx, the best WRAP design was also able to render CjPglB water-soluble without the need for detergents and with retention of function, indicating that there are multiple solutions to the water-soluble protein design problem. Taken together, our results demonstrate two promising platforms—SIMPLEx and WRAPs—for unlocking structure-function relationships of GT-C fold enzymes as well as expanding glycosylation capabilities in synthetic glycobiology.

Results

Engineering water-soluble CjPglB variants using SIMPLEx technology. We initially focused our attention on CjPglB, a GT-C fold family member consisting of 13 TMDs (**Supplementary Fig. 1a**). To render CjPglB water soluble, we designed a tripartite chimera based on the original SIMPLEx architecture^{16, 19} in which the N-terminus of CjPglB was genetically fused to a water-soluble “decoy” protein, namely ΔspMBP (*E. coli* maltose-binding protein lacking its N-terminal signal peptide), and the C-terminus was fused to an amphipathic “shield” protein, namely ApoA1* (truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain) (**Fig. 1a**). However, following expression in *E. coli* BL21(DE3) cells, the ΔspMBP-CjPglB-ApoA1* chimera (hereafter Sx-CjPglB.v1) accumulated almost entirely in the detergent-solubilized membrane fraction with virtually no detectable accumulation in the soluble fraction (**Supplementary Fig. 1b**). In fact, the expression pattern for Sx-CjPglB.v1 was similar to that observed for wild-type (wt) CjPglB without any fusion partners (**Fig. 1b**), indicating that the original SIMPLEx design was incapable of solubilizing the OST. These results were in stark contrast to what was seen for SIMPLEx-reformatted variants of human ST6Gal1 and FUT8 from the GT-A and GT-B superfamilies, respectively, which both accumulated predominantly in the soluble fraction (**Supplementary Fig. 1b**), as we had seen previously¹⁶.

We suspected that the lack of soluble expression was due to the complex topology of CjPglB. Specifically, CjPglB possesses a globular periplasmic domain that comprises the C-terminal half of the enzyme and accounts for ~40% of its mass. In the original Sx-CjPglB.v1 design, this large, extramembranous domain sits between the solubilizing ApoA1* domain and the 13 hydrophobic TMDs of CjPglB. Such an arrangement is likely suboptimal for promoting intimate contact between ApoA1* and the TMDs of the target IMP, which is a key feature of SIMPLEx-mediated solubilization¹⁹. Therefore, we redesigned the SIMPLEx architecture by moving the ApoA1* domain to the N-terminal side of CjPglB (**Fig. 1a**), which we hypothesized would provide ApoA1* with less obstructed access to the TMDs of CjPglB. When the newly redesigned ΔspMBP-ApoA1*-CjPglB construct (hereafter Sx-CjPglB.v2) was expressed in BL21(DE3) cells, we readily detected soluble expression in the cytoplasm without the need for added detergents (**Fig. 1b**), indicating that CjPglB had been converted into a water-soluble conformation. Omission of the decoy protein

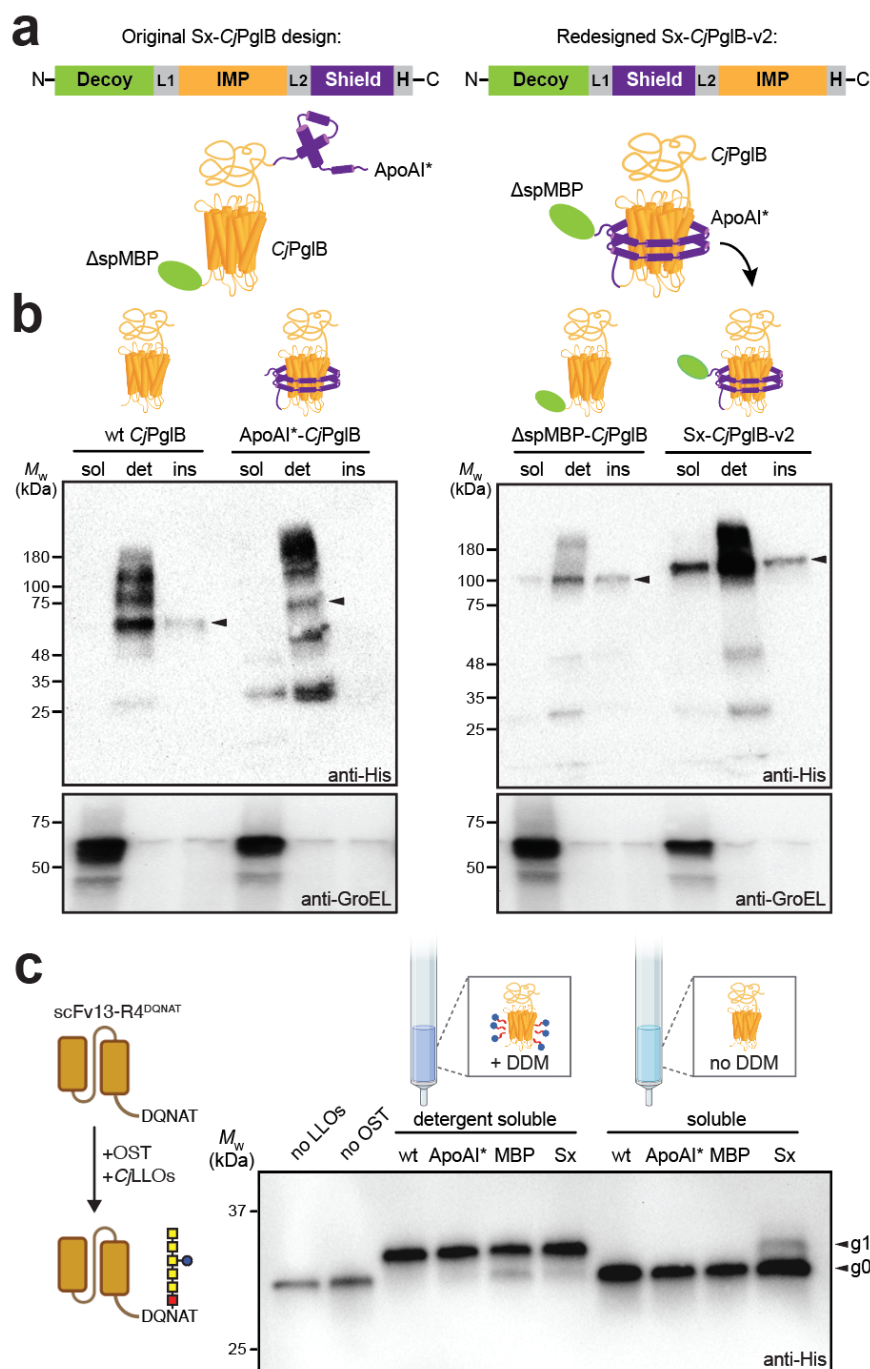


Figure 1. Redesigned SIMPLEX architecture enables solubilization of CjPglB. (a) Schematic of original and redesigned SIMPLEX architectures (see text for details). L1 and L2: two-residue linkers created by SacI and NcoI restriction sites, respectively; H: 10xHis tag. (b) Immunoblot analysis of soluble (sol), detergent solubilized (det), and insoluble (ins) fractions prepared from *E. coli* BL21(DE3) cells expressing each construct. An equivalent amount of total protein was loaded in each lane. Blots were probed with anti-His antibody to detect each of the expressed constructs and anti-GroEL antibody to confirm quality of fractionations and equivalent loading of soluble fractions. Blot on left required much longer exposure time than blot on right. (c) Immunoblot analysis of IVG reactions in which different CjPglB constructs were purified from soluble (sol) or detergent solubilized (det) fractions of BL21(DE3) cells and incubated with organic solvent-extracted CjLLOs and purified scFv13-R4^{DQNAT} acceptor protein. IVG reactions performed without CjLLOs or CjPglB served as negative controls. Blots were probed with anti-His to detect the acceptor protein. Black arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of scFv13-R4^{DQNAT}. For all immunoblots, molecular weight (M_w) markers are indicated at left and results are representative of biological replicates ($n = 3$).

resulted in no detectable accumulation in the soluble fraction, with the decoy-less ApoAI*-CjPglB construct partitioning almost exclusively to the detergent-solubilized membrane fraction in a manner that was identical to unfused wt CjPglB. It is noteworthy that expression of these two constructs was significantly lower than Sx-CjPglB.v2, requiring much longer exposure of immunoblots to generate equivalent band intensities. Omission of ApoAI* resulted in modest levels of expression in all three fractions with just a small portion accumulating in the soluble fraction that likely corresponded to soluble aggregates of CjPglB, which was seen for a previously tested decoy-IMP fusion¹⁹. Overall, our results agree with previous findings^{16, 19} that both the decoy protein and amphipathic shield protein are equally important for effective IMP solubilization. In addition to solubility, we also investigated purification yield given the strong expression of Sx-CjPglB.v2 relative to the other constructs. Following expression and purification from BL21(DE3) cells, the yield of Sx-CjPglB.v2 reached 5.1 mg/L (**Supplementary Fig. 2**), which aligned closely with the 5–10 mg/L yield values reported previously for other SIMPLEX-reformatted IMPs^{16, 19}. Conversely, the yield of wt CjPglB was only 0.5 mg/L, revealing that the SIMPLEX strategy not only promoted solubilization of CjPglB but also increased its purification yield by an order of magnitude compared to the unfused enzyme.

We next determined whether the water-soluble Sx-CjPglB.v2 construct retained enzymatic activity. To this end, we implemented an established *in vitro* glycosylation (IVG) assay²⁶, which involved purifying each CjPglB construct from the different cell fractions and subsequently combining the purified enzymes with their requisite glycosylation substrates. Specifically, lipid-linked oligosaccharides (LLOs) bearing the *C. jejuni* heptasaccharide *N*-glycan (CjLLOs) with a mixture of diNAcBac and GlcNAc as the reducing-end sugar (i.e., GalNAc₅(Glc)diNAcBac/GlcNAc, where diNAcBac = bacillosamine) were organic solvent-extracted from glycoengineered *E. coli* and used as the donor substrate. A purified single-chain Fv antibody fragment harboring two mutations to remove putative internal glycosylation sites and a DQNAT glycosylation tag and 6xHis at its C-terminus (scFv13-R4(N34L, N77L)^{DQNAT-His6}; hereafter scFv13-R4^{DQNAT}) served as the acceptor protein substrate. To evaluate the glycosylation status of scFv13-R4^{DQNAT} in IVG reactions, immunoblot analysis was performed using an anti-His antibody to detect the formation of aglycosylated (g0) and mono-glycosylated (g1) forms of the scFv13-R4^{DQNAT} acceptor protein. According to

this analysis, productive glycosylation was achieved with Sx-CjPglB.v2 that was purified from the soluble and detergent-solubilized membrane fractions (**Fig. 1c**). Importantly, because proteins purified from the soluble portion of the lysate were not supplemented with detergent, the glycosylation activity observed for this fraction was attributed to spontaneously soluble Sx-CjPglB.v2 (i.e., SIMPLEX-reformatted CjPglB that could dissolve in water without requiring additional treatments to aid dissolution). In stark contrast, wt CjPglB, ApoA1*-CjPglB, and ΔspMBP-CjPglB only exhibited activity when purified from the detergent-solubilized membrane fraction (**Fig. 1c**), consistent with their significant accumulation in this fraction. Interestingly, the ΔspMBP-CjPglB construct purified from the soluble fraction showed little to no glycosylation activity, despite its modest accumulation in this fraction, providing further evidence that this construct was in a partially soluble but misfolded state. Collectively, these data establish that our SIMPLEX methodology is useful for solubilizing a member of the GT-C superfamily, CjPglB, while preserving its sequence, fold, and function.

Modularity of SIMPLEX enables fine-tuning of solubility and catalytic activity.

Although the Sx-CjPglB.v2 construct purified from the soluble fraction was active, its glycosylation efficiency (measured as the intensity of the g1 band divided by the sum of the intensities of the g0 and g1 bands in the immunoblot in **Fig. 1c**) was considerably lower than unfused wt CjPglB purified from the detergent-solubilized membrane fraction (22% vs 100% efficient, respectively) under the conditions tested. Therefore, we sought to enhance the efficiency of the water-soluble CjPglB construct by leveraging the modularity of the SIMPLEX architecture. Specifically, we generated a panel of variants in which the size of the decoy domain was decreased while the size of the amphipathic shield domain was increased (**Fig. 2a**). In the case of the decoy, we replaced the ~42-kDa ΔspMBP domain with smaller hydrophilic sequences, namely small ubiquitin-like modifier protein (SUMO; ~11 kDa) or a triply repeated FLAG epitope (DYKDHDGDYKDHDIDYKDDDDK; ~1 kDa). We chose these domains because their high intrinsic solubility and compact size are well known to enhance expression and purification of their fusion partners without impacting activity^{27, 28}. In the case of the shield, we incrementally increased the size of the ApoA1* domain, which naturally contains 10 amphipathic helices (10H)²⁹. Specifically, head-to-tail fusion of ApoA1* with the C-terminal half of ApoA1* containing helices 6–10 resulted in a 15-helix (15H) design while fusing an additional full-length copy of ApoA1* containing

helices 1–10 resulted in a 20-helix (20H) design. The rationale for lengthening this domain was based on our previous low-resolution structural analysis, which showed that ApoA1* tended to wrap around the transmembrane helices of its IMP fusion partner¹⁹, consistent with the “belt”-like conformation it adopts with phospholipids³⁰. To shield the 13 TMDs of CjPglB, which present considerably more hydrophobic surface area than any of the previously solubilized IMPs (which possessed between 3–7 TMDs), we speculated that a longer belt might be needed.

Immunoblot analysis revealed that all newly constructed variants promoted soluble expression of CjPglB without the need for detergents, with similar accumulation patterns as the Sx-CjPglB.v2 construct (**Fig. 2b**). Overall, the larger decoys combined with shorter shields resulted in the highest levels of soluble expression as exemplified by the ΔspMBP and SUMO decoys fused with the 10-helix (10H) ApoA1* domain. It is also worth noting that whereas all fusions involving the ΔspMBP decoy were stably expressed with little visible degradation, the FLAG and SUMO constructs exhibited greater amounts of degradation. Importantly, when the soluble fractions were tested for IVG activity, each construct was observed to catalyze glycan installation on the scFv13-R4^{DQNA}T acceptor protein albeit to varying extents (**Fig. 2c**). Unexpectedly, the constructs that exhibited the most significant solubilization, such as ΔspMBP-10H-CjPglB (a.k.a. Sx-CjPglB.v2) and SUMO-10H-CjPglB that consisted of larger decoys fused with the shortest shield, were the least active in terms of glycosylation efficiency (all <60% efficient). On the other hand, designs involving shorter decoys fused to longer shields were among the most active, with 3xFLAG-15H-CjPglB showing ~80% glycosylation efficiency.

We suspected that the lower glycosylation efficiency seen for the ΔspMBP-based constructs might be due to interference caused by the relatively bulky ΔspMBP domain. To test this notion, we attempted to create greater spacing and flexibility between the domains of the least active ΔspMBP-15H-CjPglB construct. To this end, we replaced the two-residue linkers (EL and MW) at the decoy:shield and shield:CjPglB junctions with alternative 27-residue linkers, specifically a flexible GlySer linker ([GSG]₉) or a more rigid TEV protease recognition site-based linker (GSG[ENLYFQG]₃GSG) (**Fig. 2a**). When either linker was placed between the ΔspMBP decoy and 15H shield, there was only modest improvement in glycosylation efficiency for the GlySer linker and none for the TEV-based linker (**Fig. 2d**). However,

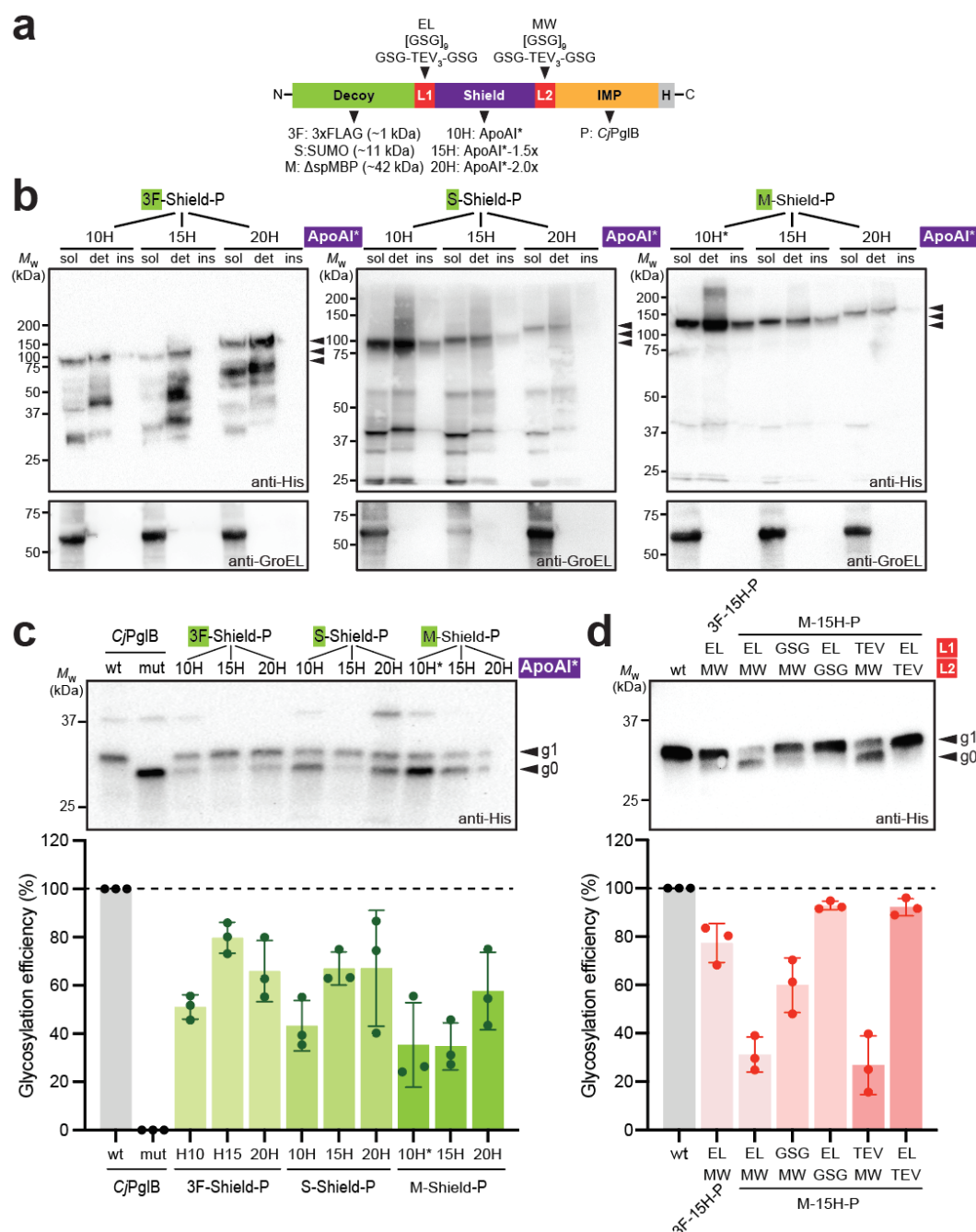


Figure 2. Discovery of highly soluble and highly active SIMPLEX-reformatted C/PglB variants. (a) Schematic of SIMPLEX modularity depicting interchangeable decoy, shield, and linker domains tested in this work. EL: glutamic acid-leucine; MW: methionine-tryptophan; GSG: glycine-serine-glycine linker; TEV: ENLYFQG protease recognition sequence. (b) Immunoblot analysis of soluble (sol), detergent solubilized (det), and insoluble (ins) fractions prepared from BL21(DE3) cells expressing each of the indicated SIMPLEX constructs. An equivalent amount of total protein was loaded in each lane. Blots were probed with anti-His to detect each of the expressed constructs and anti-GroEL antibody to confirm quality of fractionations and equivalent loading of soluble fractions. Black arrows denote full-length expression products. (c) Immunoblot analysis of IVG reactions (top) in which purified scFv13-R4^{DQNAT} acceptor protein was incubated with organic solvent-extracted CjLLOs and SIMPLEX constructs in (b) that were purified from soluble (sol) or detergent solubilized (det) fractions of BL21(DE3) cells. Reactions involving detergent-solubilized wt C/PglB enzyme lacking fusion partners served as positive control (lane 1) while reactions involving detergent-solubilized, catalytically inactive mutant (mut) C/PglB enzyme served as negative control (lane 2). Blots were probed with anti-His to detect the acceptor protein. Black arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of scFv13-R4^{DQNAT}. (d) Immunoblot analysis of IVG reactions (top) in which purified scFv13-R4^{DQNAT} acceptor protein was incubated with organic solvent-extracted CjLLOs and SIMPLEX constructs in (b) that were purified from soluble (sol) or detergent solubilized (det) fractions of BL21(DE3) cells. Reactions involving detergent-solubilized wt C/PglB enzyme lacking fusion partners served as positive control (lane 1) while reactions involving detergent-solubilized, catalytically inactive mutant (mut) C/PglB enzyme served as negative control (lane 2). Blots were probed with anti-His to detect the acceptor protein. Black arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of scFv13-R4^{DQNAT}.

For all immunoblots in (b) and (c), molecular weight (M_w) markers are indicated at left and results are representative of biological replicates ($n = 3$). Glycosylation efficiency (bottom) was determined based on densitometric quantification of the fraction of glycosylated acceptor protein expressed as the ratio $g1/[g0+g1]$. Data are mean of biological replicates ($n = 3$) $\pm$ SD. (d) IVG reactions as in (c) but with SIMPLEX constructs based on 3xFLAG or Δ spMBP decoys fused to ApoA1*-1.5x (15H) with different linker sequences as indicated. IVG reaction involving detergent-solubilized wt CjPglB without any fusions served as positive control (lane 1).

when the same linkers were positioned between the 15H shield and CjPglB enzyme, both constructs exhibited a dramatic increase in glycosylation efficiency that reached near 100% and rivaled that of the detergent-solubilized wt CjPglB without any fusion partners. This high glycosylation activity was achieved without compromising the robust soluble expression that was characteristic of the much less active Sx-CjPglB.v2 construct. In fact, the optimized Δ spMBP-EL-15H-[GSG]₉-CjPglB construct (hereafter Sx-CjPglB.v3) reached 5.6 mg/L when purified directly from the soluble fraction, which was similar to the yield of Sx-CjPglB.v2 purified from the same fraction and >10-fold greater than the yield of wt CjPglB purified from the detergent-solubilized membrane fraction (**Supplementary Fig. 2**). Using the Sx-CjPglB.v3 that was purified from the soluble fraction in IVG reactions resulted in highly efficient glycosylation of the scFv13-R4^{DQNAT} acceptor protein without the need for added detergent, whereas wt CjPglB derived from this same fraction showed no measurable activity (**Supplementary Fig. 3**). Taken together, these results reveal how the straightforward interchangeability of the decoy and shield domains can be used to fine-tune the solubility and activity of SIMPLEX-reformatted IMPs, leading to the identification of catalytically robust designs as exemplified by Sx-CjPglB.v3.

Machine learning-based design of water-soluble CjPglB variants. In contrast to the universal solubilization afforded by the SIMPLEX methodology, *de novo* designed protein WRAPs (water-soluble RFdiffused amphipathic proteins) are an alternative deep learning-based design approach for custom solubilization of IMP targets ²⁵. Similar to ApoA1, WRAPs have a polar exterior and a nonpolar interior complementary to the lipid-facing hydrophobic surface of the target. However, in a notable departure from the SIMPLEX method, each WRAP is computationally tailored to surround the lipid-interacting hydrophobic surfaces of its IMP target using RFdiffusion ³¹ and ProteinMPPN ³², generative models for *de novo* protein backbone and sequence design, respectively. Here, we used a similar approach to generate 20 alpha-helical, CjPglB-specific WRAP domains that were designed to render the OST stable and

water-soluble without the need for detergents (**Supplementary Fig. 4**). The four highest scoring designs based on pLDDT binder scores (**Supplementary Table 1**) were each genetically fused to the N-terminus of CjPglB with a flexible GlySer linker ([GGGS]₃) (**Fig. 3a**) and evaluated for soluble expression in the cytoplasm of *E. coli* BL21(DE3) cells. Following purification of the helical WRAP-CjPglB fusion proteins from the soluble fraction of whole-cell lysates, we determined that two of the designs, WRAP-CjPglB.d1 and WRAP-CjPglB.d3, were soluble (**Supplementary Fig. 5**).

Because WRAP-CjPglB.d3 was only weakly detected in the soluble fraction, we turned our attention to WRAP-CjPglB.d1, which showed substantially greater soluble expression. To benchmark the solubility of the WRAP construct, we performed immunoblot analysis of fractionated lysates derived from *E. coli* BL21(DE3) cells expressing either WRAP-CjPglB.d1, unfused wt CjPglB or Sx-CjPglB.v3. This analysis revealed that WRAP-CjPglB.d1 accumulated in the soluble fraction to an extent that greatly exceeded that of Sx-CjPglB.v3 (**Fig. 3b**), demonstrating that the WRAP strategy was highly effective at rendering the 13-TMD OST water-soluble in the cytoplasm without the need for detergents. As before, unfused wt CjPglB partitioned exclusively in the detergent-solubilized membrane fraction. In terms of yield, we recovered 3.3 mg/L of WRAP-CjPglB.d1, which was 6-fold greater than the yield of unfused wt CjPglB but less than the yields for the SIMPLEXed CjPglB constructs (**Supplementary Fig. 2**). This latter difference was somewhat surprising given the much stronger expression of the WRAPed construct but could relate to differences in how the enzymes were purified.

To investigate whether WRAPed CjPglB retained the enzymatic function of the native OST, we again turned to the IVG assay using scFv13-R4^{DQNAT} as acceptor protein substrate. Importantly, clear evidence of glycosylation was observed for WRAP-CjPglB.d1 that was purified from the soluble and detergent-solubilized membrane fractions (**Fig. 3c**). This activity was on par with that measured for Sx-CjPglB.v3, with both water-soluble enzymes exhibiting activity that was 60-80% of the activity measured for detergent-solubilized wt CjPglB (**Fig. 3d**). These data establish that our deep learning-based design approach can generate custom alpha-helical WRAP designs that complement the more universal SIMPLEX designs in their ability to create soluble, functionally active versions of CjPglB.

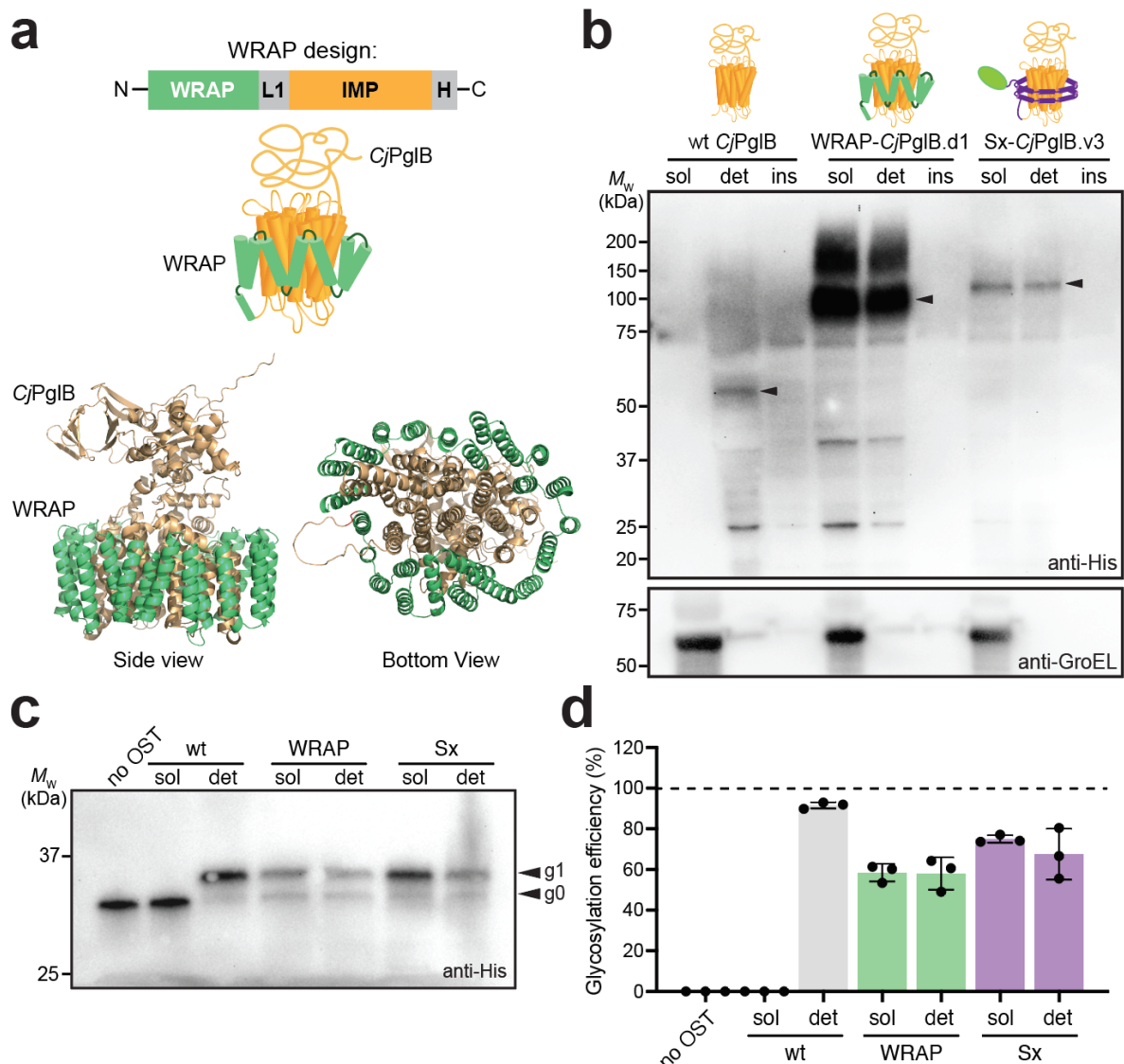


Figure 3. Solubilization of CjPgIB using computationally designed helical WRAP domain. (a) Schematic representation (top) and design model (bottom) of WRAP-CjPgIB.d1 construct with helical WRAP solubilizing domain (green) custom designed to be complementary to the lipid-facing hydrophobic surface of CjPgIB (orange). L1: flexible (GGGS)₃ linker; H: 10xHis tag. (b) Immunoblot analysis of soluble (sol), detergent solubilized (det), and insoluble fractions (ins) prepared from BL21(DE3) strains expressing wt CjPgIB, WRAP-CjPgIB.d1, and Sx-CjPgIB.v3. Blot was probed with anti-His to detect each of the expressed constructs. An equivalent amount of total protein was loaded in each lane for wt CjPgIB and Sx-CjPgIB.v3; three times less total protein was loaded for WRAP-CjPgIB.d1 samples due to its much higher expression level relative to the other constructs. Black arrows denote full-length expression products. (c) Immunoblot analysis of IVG reactions in which purified scFv13-R4^{DQ_{NAT}} acceptor protein was incubated with organic solvent-extracted CjLLOs and CjPgIB constructs in (b) that were purified from soluble (sol) or detergent solubilized (det) fractions of BL21(DE3) cells. Reactions involving detergent-solubilized wt CjPgIB enzyme lacking fusion partners served as positive control (lane 1) while reactions involving detergent-solubilized, catalytically inactive mutant (mut) CjPgIB enzyme served as negative control (lane 2). Blots were probed with anti-His to detect the acceptor protein. Black arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of scFv13-R4^{DQ_{NAT}}. Molecular weight (M_w) marker is indicated at left and results are representative of biological replicates ($n = 3$). (d) Glycosylation efficiency was determined based on densitometric quantification of the fraction of glycosylated acceptor protein expressed as the ratio $g1/[g0+g1]$. Data are mean of biological replicates ($n = 3$) $\pm$ SD.

Kinetic analysis of water-soluble CjPglB variants. Having established the water solubility and glycosylation function of the SIMPLExed and WRAPed versions of CjPglB, we next sought to more rigorously quantify their catalytic performance. To this end, we utilized a peptide substrate-based IVG assay that has been used previously by us and others to establish reaction kinetics for CjPglB and a few of its homologs³³⁻³⁶. This assay is performed with the same solvent-extracted CjLLOs as glycan substrate donor together with a short acceptor peptide substrate containing the DQNAT glycosylation sequon that is terminally labeled with a tetramethylrhodamine fluorophore (5-TAMRA-GSDDQNATF-NH₂, where the underlined region represents the sequon) (**Fig. 4a**). The extent of TAMRA peptide glycosylation in these reactions is tracked using in-gel fluorescence analysis whereby the increase in molecular weight corresponding to the addition of the ~1-kDa heptasaccharide is visualized directly in tricine-SDS-PAGE gels. Incubation of these substrates with unfused wt CjPglB, Sx-CjPglB.v3, or WRAP-CjPglB.d1 resulted in the formation of fluorophore-labeled glycopeptide products dependent on the fraction from which the OST was purified (**Supplementary Fig. 6a**). Specifically, whereas all three enzymes glycosylated the TAMRA-peptide when the enzymes were purified from the detergent-solubilized membrane fraction, only the Sx-CjPglB.v3 and WRAP-CjPglB.d1 enzymes were able to catalyze TAMRA peptide glycosylation when the enzymes were purified from the soluble fraction, consistent with our IVG results using scFv13-R4^{DQNAT} as acceptor substrate. These water-soluble OSTs were determined to be 85% and 30% efficient, respectively, while unfused wt CjPglB purified from the detergent-solubilized membrane fraction was 91% efficient (**Supplementary Fig. 6b**).

Using this TAMRA peptide glycosylation assay, we proceeded to estimate the kinetic constants for all three OSTs. Specifically, Michaelis-Menten kinetics were determined by performing IVG reactions with increasing concentrations of the TAMRA peptide substrate and quantifying the in-gel fluorescence intensities of the aglycosylated (g0) and glycosylated (g1) bands as a function of peptide concentration (**Fig. 4b**) and used these data to determine K_m and V_{max} values (**Table 1**). From this analysis, the water-soluble Sx-CjPglB.v3 and detergent-solubilized wt CjPglB constructs exhibited comparable kinetics, clearly outperforming the WRAP-CjPglB.d1 construct (**Table 1**). One possible explanation for this discrepancy is that WRAPs are designed to bind tightly to their IMP target, which in turn could affect their catalytic

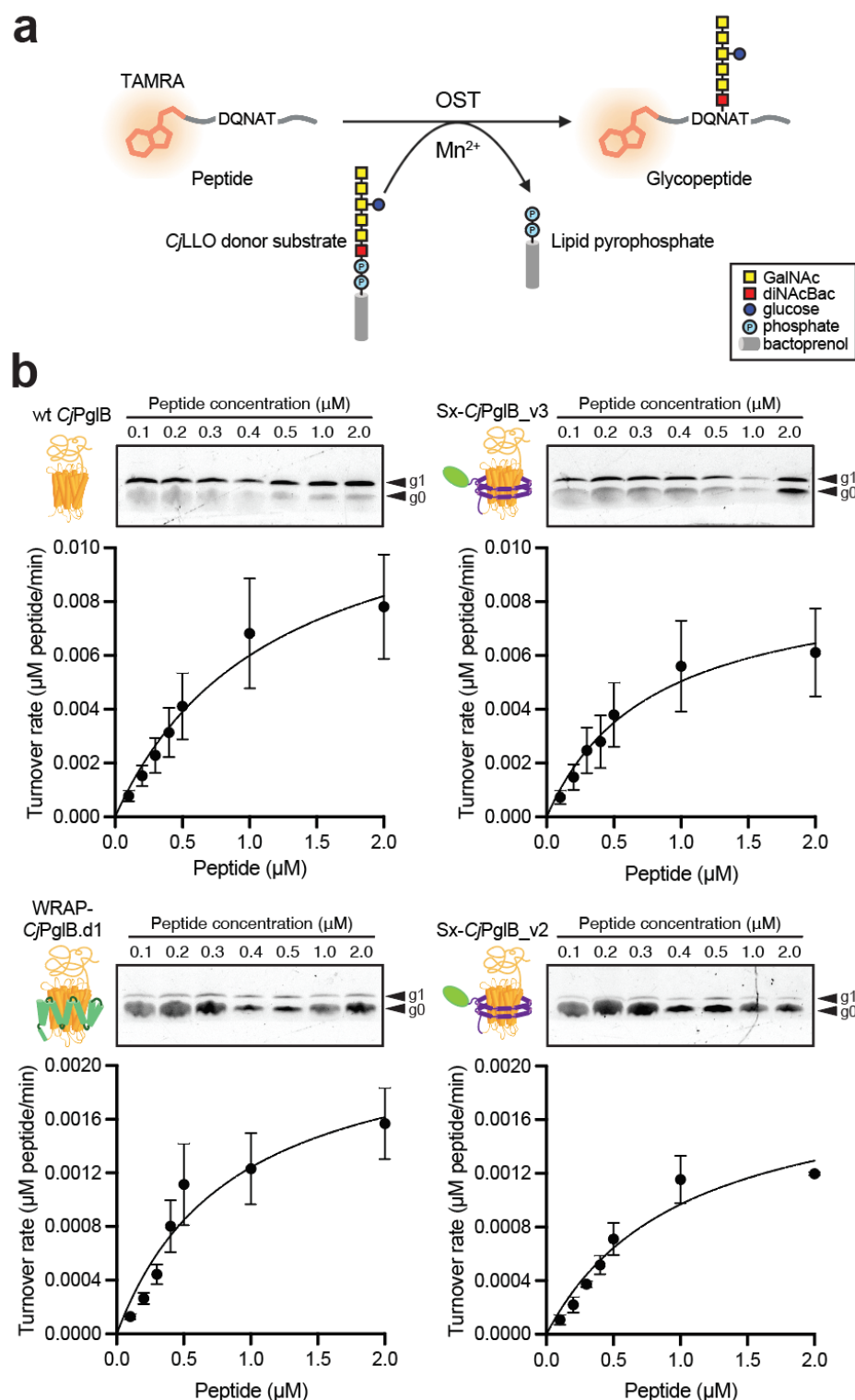
activity. However, we also note that whereas Sx-CjPglB.v3 underwent significant optimization of its linker and decoy domains to improve its activity, the WRAP-CjPglB.d1 construct was not subjected to any such optimization. Thus, a fairer comparison is between the straight out-of-the-box WRAP-CjPglB.d1 construct and its non-optimized Sx-CjPglB.v2 counterpart. Indeed, the kinetics of these two constructs was similar, with WRAP-CjPglB.d1 slightly outperforming Sx-CjPglB.v2 (**Figure 4b** and **Table 1**).

We next decided to determine and compare initial turnover rates because impractically high peptide concentrations would be required to reach $V_{\max}$, as was noted previously³³. Specifically, the turnover rate for all four OSTs was determined from the initial, linear range of time course data (**Supplementary Fig. 7**). This analysis revealed a similar trend in terms of catalytic performance with the turnover rates decreasing in the order: wt CjPglB > Sx-CjPglB.v3 > WRAP-CjPglB.d1 > Sx-CjPglB.v2. Collectively, these kinetic data help to explain the differences in glycosylation efficiency achieved by these enzymes in the context of the scFv13-R4^{DQNAT} acceptor substrate while also providing further confirmation that our water-soluble CjPglB designs retained the catalytic function of the native OST without the need for detergents. It is also worth noting that the kinetic constants determined here compared favorably with values reported previously for detergent-solubilized wt CjPglB assayed with LLOs derived from either chemical synthesis or cellular biosynthesis^{34, 35}.

Table 1. Kinetic parameters for wild-type and water-soluble CjPglB enzymes

OST	K_M (μM)	$V_{\max}$ (peptide h^{-1})	Turnover rate (h^{-1})
wt CjPglB	1.2 ± 0.18	0.78 ± 0.26	0.072 ± 0.001
Sx-CjPglB.v2	1.0 ± 0.24	0.12 ± 0.01	0.023 ± 0.002
Sx-CjPglB.v3	0.78 ± 0.14	0.54 ± 0.16	0.056 ± 0.003
WRAP-CjPglB.d1	0.87 ± 0.20	0.14 ± 0.03	0.044 ± 0.006

Kinetic constants (K_M and $V_{\max}$) were calculated from IVG data in Figure 4 while turnover rate was calculated from IVG data in Supp. Fig. 7. Reactions were performed with extracted CjLLOs and fluorescent 5-TAMRA-GSDQNATF-NH₂ as substrates. Data are the average of biological replicates ($n = 3$) $\pm$ SD.



DISCUSSION

In this work, we demonstrate that even the most topologically complex GTs—those belonging to the GT-C structural superfamily—can be systematically converted into functional, water-soluble enzymes using rational and computational protein engineering approaches. By applying both SIMPLEX, a universal solubilization platform, and WRAPs, a custom deep learning–guided design strategy, we overcame the longstanding challenge of producing multi-pass membrane glycosyltransferases in a tractable, detergent-free format. The successful reformatting of CjPglB, an archetypal GT-C enzyme containing 13 transmembrane helices, highlights the robustness of these platforms and provides a blueprint for solubilizing other GT-C family members whose structures and biochemical properties remain poorly understood.

Importantly, both strategies produced water-soluble PglB variants that retained the defining catalytic hallmarks of wild-type enzyme function. SIMPLEX-solubilized CjPglB performed with kinetic parameters and glycosylation efficiencies comparable to detergent-solubilized wild type, and WRAP-designed variants achieved similar functional outcomes despite being built from entirely distinct protein design principles. The convergence of function across such divergent engineering approaches underscores a key insight: many of the topological requirements for membrane integration can be decoupled from the catalytic machinery itself, provided that the hydrophobic surfaces of the enzyme are adequately shielded and its dynamic architecture is preserved.

Together, these results establish SIMPLEX and WRAPs as versatile and complementary strategies for overcoming the expression and solubility bottlenecks that have constrained molecular characterization of the GT-C superfamily. While SIMPLEX is universally applicable, WRAPs require custom protein design, which in turn contributes to stability and enables structure determination. As structural data for GT-C enzymes remain scarce, the ability to generate high-yield, monodisperse, detergent-free variants creates new opportunities for biochemical, biophysical, and structural interrogation, including crystallography, cryo-EM, and deep mutational scanning. Moreover, these approaches could be extended beyond GT-Cs to other polytopic integral membrane enzymes whose inherent hydrophobicity currently limits mechanistic study and biotechnological deployment.

Despite these advances, several open questions remain. Future efforts will be needed to determine how solubilization affects long-range conformational dynamics of GT-C enzymes, particularly those linked to membrane-dependent substrate binding or conformational gating. Additionally, while both design approaches preserved catalytic activity toward model substrates, broader analyses will be required to assess how well substrate specificity, glycan preferences, or in vivo selectivity are maintained across diverse donor and acceptor contexts. Finally, the full generalizability of these solubilization strategies across the phylogenetic diversity of GT-C enzymes—including eukaryotic STT3 paralogs—remains an exciting direction for future exploration.

Overall, our work provides foundational evidence that structurally complex membrane-integrated glycosyltransferases can be reformatted into stable, active, water-soluble catalysts. By enabling high-yield expression, functional characterization, and integration into cell-free and synthetic glycosylation platforms, SIMPLEx and WRAPs open new avenues for discovery and application across glycobiology, enzymology, and glycoengineering.

MATERIALS AND METHODS

Bacterial strains and plasmids. *E. coli* strain NEB5 α (New England Biolabs), a derivative of DH5 α , was used for all plasmid construction, maintenance, and DNA purification. *E. coli* strain BL21(DE3) (Novagen) was used for all protein expression experiments. Plasmid pSN18³⁷ was used for expression of wt CjPglB, while plasmid pET28a-scFv13-R4(N34L, N77L)^{DQNAT-His6}³⁶ was used for expression of the scFv13-R4^{DQNAT} acceptor protein. *E. coli* strain CLM24 carrying plasmid pMW07-pgl Δ B³⁸, which encodes the entire *C. jejuni* *pgl* pathway except for CjPglB³⁹, was used for preparing CjLLOs. CLM24 is a W3110 derivative with a deletion in the O-antigen ligase WaaL that prevents glycan transfer to lipid A-core⁴⁰, enabling accumulation of the *C. jejuni* heptasaccharide *N*-glycan on undecaprenyl pyrophosphate (Und-PP).

To express CjPglB in the original SIMPLEx format, plasmid pET28a-Sx-CjPglB.v1 was constructed by PCR amplifying the gene encoding CjPglB from pSN18 and inserting it between Δ spMBP and ApoAI* in plasmid pET28(+)- Δ spMBP-(*Nde*I)-GT-(*Eco*RI)-ApoAI*¹⁶ using Gibson Assembly⁴¹. To express CjPglB in the redesigned SIMPLEx format, plasmid pBAD-Sx-CjPglB.v2 encoding *E. coli* MBP, human ApoAI*, and CjPglB in the pSN18 backbone was constructed commercially (Integrated DNA

Technologies). The resulting plasmid placed the Sx-CjPglB.v2 construct under control of an arabinose-inducible promoter and introduced unique restriction sites flanking each of the individual genes and a C-terminal 10xHis tag as follows: pBAD-(*SpeI*)- Δ spMBP-(*SacI*)-ApoAI*-(*NcoI*)-CjPglB-(*EcoRI*)-10xHis. The plasmids pBAD- Δ spMBP-CjPglB and pBAD-ApoAI*-CjPglB for expressing control constructs lacking either the ApoAI* or Δ spMBP domains, respectively, were constructed by PCR amplification of the complete pBAD-Sx-CjPglB.v2 plasmid excluding either the ApoAI* or Δ spMBP sequence, followed by self-ligation of the linear PCR product in *E. coli* NEB5 α cells. To generate SIMPLEX variants with alternative decoy domains, DNA encoding 3xFLAG and SUMO domains was synthesized (Integrated DNA Technologies), amplified by PCR, and subsequently used to replace the DNA encoding the Δ spMBP domain in pBAD-Sx-CjPglB.v2 by Gibson Assembly. Likewise, to introduce longer ApoAI* domains consisting of 15 and 20 helices, DNA encoding the 15H and 20H helices was synthesized (Integrated DNA Technologies), PCR amplified, and used to replace the 10-helix (10H) ApoAI* domain by Gibson Assembly. To introduce alternative linkers, DNA encoding the flexible GlySer linker ([GSG]₉) and the TEV protease recognition site-based linker (GSG[ENLYFQG]₃GSG) were synthesized (Integrated DNA Technologies), PCR amplified and ligated into plasmid pBAD-Sx-CjPglB.v2 at either the *SacI* or *NcoI* restriction site. To express the WRAP-CjPglB designs, DNA encoding the top four WRAP designs was codon optimized for expression in *E. coli* and synthesized (Twist Bioscience). The resulting synthetic DNA was PCR amplified and inserted upstream of the gene encoding wt CjPglB in pSN18 by Gibson Assembly, resulting in plasmids pBAD-WRAP-CjPglB.d1 to v4. All newly constructed plasmids were generated using standard molecular cloning techniques and verified by Oxford nanopore whole plasmid sequencing (Eurofins Genomics) or Sanger sequencing at the Genomics Facility of the Cornell Biotechnology Resource Center.

Design of protein WRAPs. Initial WRAPs were generated parametrically using sushimaki (<https://github.com/davidekim/sushimaki>), a python script that uses DeepTMHMM⁴² to predict the membrane spanning residues and topology of the target used to estimate the radius (*tr*), height (*th*), and central axis of the transmembrane region, and PyRosetta⁴³ to generate and place the helical WRAP around the transmembrane region of the target. The main purpose of the script is to roughly

generate WRAPs around a given target to use as inputs for structure refinement using partial RFdiffusion^{31, 44}. Ideal helices of hn residues per helix are generated, where hn is determined by th , and the number of helices (n) is determined by tr using the following equations, respectively:

$$hn = \text{round}(((th + 5)/5.4) * 3.6)$$

$$n = \text{round}(\pi * 2 * (tr + 6)/7)$$

hn is calculated based on the rise per residue of an ideal helix, 5.4/3.6, plus a 5-angstrom buffer added to th . n is determined from the circumference of the WRAP divided by 7, a value that was empirically derived from WRAPs in a previous study²⁵. hn and n can also be provided as command line options where th and tr are calculated using the same equations rather than estimated by DeepTMHMM. To generate the WRAP, sushimaki places a helix approximately along the transmembrane central axis by superimposing the center of masses of the top and bottom 4 residue CA atoms and all CAs of the helix to the center of masses of the top and bottom 2 residue CAs of each membrane spanning chain segment and all spanning CAs of the transmembrane region, respectively. The aligned helix is then translated perpendicular to the transmembrane central axis at a distance of tr plus 6 angstroms to roughly accommodate side chain packing. The helix up and down orientation is flipped if the N or C terminus is on the opposite side of the C or N terminus of the target, respectively. The helix is then rotated around the central axis in 5 degree increments until the distance between the respective termini is at the minimum to enable fusion of the WRAP to the target with a short flexible linker. The remaining helices are generated as copies of the helix and then placed evenly around the transmembrane region by rotating around the central axis with alternating up and down orientations. The helices are then connected by 3 residue loops into a single chain using the PyRosetta loop modeler. Both forward and reverse rotating WRAPs with -30, 0, and 30 degree rotation offsets are generated for N and C terminal target fusions for a total of 12 WRAPs per n and hn parameter values used. For the target protein, 5OGL, WRAPs were generated using n values of 19 to 24 helices and hn values of 16 and 17 residues to sample a larger range of WRAP structures verses $n = 19$ and $hn = 17$ derived from DeepTMHMM for 5OGL.

The WRAPed target complexes were then refined using partial RFdiffusion conditioned on the target structure with “diffuser.partial_T” of 30 and “diffuser.T” of 50

steps total to generate 1440 designs. These partially diffused backbones were then sequence designed and validated by a protein-protein interaction design pipeline described previously⁴⁵ involving SolubleMPNN³², Rosetta FastRelax⁴⁶, and AlphaFold2 (AF2)⁴⁷ initial guess. Briefly described here, the pipeline uses SolubleMPNN to design sequences for the partially diffused WRAP backbones while keeping the target sequence fixed. These designs are also Rosetta FastRelaxed and sequence re-designed. The structures of the designed complexes are then predicted using AF2 initial guess. 8 designs were selected with binder (WRAP) pLDDT > 80, interface PAE (PAE_i) < 10, and SAP (spatial aggregation propensity)⁴⁸ < 600 from 2880 AF2 models. To further improve these metrics, an iterative partial diffusion optimization strategy (https://github.com/davidekim/ppi_iterative_opt) was used for each of the 8 selected designs involving iterations of partial diffusion, SolubleMPNN plus Rosetta FastRelax, and AF2 validation, where the design with the lowest PAE_i in each iteration is used for the next iteration. Up to 10 iterations were run, and 5 final designs with the highest pLDDTs were selected from the top 20 designs with the lowest PAE_i values from 13,053 designs. 4 designs with exposed hydrophobic transmembrane target residues were excluded upon visual inspection.

Subcellular fractionation. For recombinant expression of wt CjPglB and the water-soluble CjPglB variants, *E. coli* BL21(DE3) cells harboring the appropriate expression plasmid were cultured overnight at 37 °C in 10 mL Luria Bertani (LB) medium supplemented with 100 µg/mL ampicillin. Overnight cells were subcultured into 1 L of Terrific Broth (TB) containing the same antibiotic and grown at 37 °C with shaking until the optical density at 600 nm (OD₆₀₀) reached ≈ 0.7. Protein expression was induced by addition of L-arabinose to a final concentration of 0.2% (v/v), and cultures were incubated at 16 °C overnight (~16–20 h). Cells were harvested by centrifugation at 5,000 ×g for 15 min at 4 °C and resuspended in Buffer A (50 mM HEPES, 250 mM NaCl, pH 7.4; sterile-filtered). Cell lysis was performed using an EmulsiFlex-C5 high-pressure homogenizer (Avestin). The lysates were cleared by centrifugation at 3,000 ×g for 20 min at 4 °C to remove cellular debris. The clarified lysates were subjected to ultracentrifugation at 100,000 ×g for 2 h at 4 °C to fractionate soluble and membrane-associated proteins. The supernatant from the ultracentrifugation step was collected as the soluble fraction. The resulting pellet was resuspended in Buffer B (50 mM HEPES, 250 mM NaCl, 1% (w/v) n-dodecyl-β-D-maltoside (DDM), pH 7.4; sterile-

filtered) and incubated at room temperature for 1 h to extract membrane proteins. Following detergent solubilization, samples were again centrifuged at 100,000 ×g for 1 h at 4 °C. The resulting supernatant was collected as the detergent-solubilized membrane fraction, while the remaining pellet was resuspended in Buffer B and defined as the insoluble fraction. Whole-cell lysate samples (i.e., lysates not subjected to ultracentrifugation) were retained as whole cell fractions for comparative analysis.

Protein purification. CjPglB proteins were purified from the soluble fraction of the lysate, the detergent-solubilized fraction of the lysate, or the whole-cell fraction using immobilized metal affinity chromatography (IMAC). For purification from whole-cell and soluble fractions, all buffers were prepared using Buffer A, while purification from detergent-solubilized fractions involved buffers prepared using Buffer B. Ni-NTA agarose resin (ThermoFisher) was equilibrated in the appropriate buffer by washing with five column volumes (CVs). Resin was added to the prepared fractions and incubated at 4 °C with gentle agitation. For wild-type CjPglB purified from the detergent-solubilized fraction, incubation was performed overnight at 4 °C. All other incubations were performed for 1 h at 4 °C. After binding, the resin was washed with 20 mM imidazole-containing wash buffer prepared from either Buffer A or B. Bound enzymes were eluted with elution buffer supplemented with 300 mM imidazole in the same base buffer. Elution was monitored by absorbance at 280 nm using a spectrophotometer (NanoDrop) at 1-mL intervals, and fractions were pooled until the absorbance signal approached baseline. Eluted proteins were buffer exchanged into either Buffer A or B using centrifugal concentrators (ThermoFisher) prior to downstream analysis.

For final yield determination and kinetic studies, all CjPglB proteins were further purified via either affinity chromatography using amylose resin for SIMPLEX constructs or size-exclusion chromatography (SEC) for wt and WRAPed CjPglB constructs. Amylose purification was performed by 10-fold dilution of IMAC-eluted samples into equilibration buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, pH 7.2). Pre-washed amylose resin (New England Biolabs) was added and incubated with samples for 1 h at 4 °C. After washing, bound proteins were eluted with equilibration buffer supplemented with 20 mM maltose, and dialyzed into Buffer A overnight. SEC was performed using a Superdex 200 Increase column (Cytiva) pre-equilibrated in Buffer A (for WRAP-CjPglB.d1) or Buffer B (for wt CjPglB). Eluted fractions were

collected and analyzed based on chromatographic profiles. Purity assessments were conducted by SDS-PAGE analysis followed by Coomassie staining of gels.

The scFv13-R4^{DQNAT} acceptor protein was purified from whole-cell lysates by IMAC. Briefly, BL21(DE3) cells carrying pET28a-scFv13-R4(N34L, N77L)^{DQNAT-His6} were initially grown overnight in 10 mL of LB supplemented with 50 µg/mL kanamycin, subcultured into 1 L of LB with the same antibiotic on the following day, and induced with 0.1 mM IPTG at 20 °C overnight. Cells were harvested and resuspended in Buffer C (20 mM Tris-HCl, 250 mM NaCl, pH 7.2; sterile-filtered) and lysed using an EmulsiFlex-C5 high-pressure homogenizer (Avestin). Lysates were clarified by centrifugation at 3,000 ×g for 20 min at 4 °C to remove cell debris. Ni-NTA agarose resin (ThermoFisher) was equilibrated in Buffer C and incubated with clarified lysates for 1 h at 4 °C with gentle agitation. After binding, the resin was washed with Buffer C supplemented with 20 mM imidazole with five CVs. Bound proteins were eluted with Buffer C supplemented with 300 mM imidazole. Elution was monitored at 280 nm using a spectrophotometer (NanoDrop) in 1-mL fractions, which were pooled until the absorbance signal approached baseline. Eluted proteins were buffer exchanged into Buffer C using PD-10 desalting columns (Cytiva) according to the manufacturer's instructions.

Coomassie staining and immunoblotting. Protein samples were denatured by addition of sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) loading buffer containing 10% β-mercaptoethanol, followed by heating at 90 °C for 10 min. Electrophoretic separation was performed using either precast 4–20% Tris-Glycine gels (Invitrogen) or AnyKD Mini-PROTEAN TGX gels (Bio-Rad). Gels from Bio-Rad were specifically used for scFv13-R4^{DQNAT} samples, while Invitrogen gels were used for all other protein samples. For total protein visualization, gels were stained using Coomassie Brilliant Blue G-250 (Bio-Rad) following the manufacturer's instructions. Stained gels were imaged using a standard flatbed scanner or gel documentation system. For immunoblotting, following electrophoresis, proteins were transferred from SDS-PAGE gels onto polyvinylidene difluoride (PVDF) membranes (Immobilon-P, Millipore) using a wet tank transfer system (ThermoFisher), following the manufacturer's protocol. Membranes were blocked for 1 h at room temperature or overnight at 4 °C in TBST buffer (20 mM Tris, 150 mM NaCl, 0.05% Tween-20) containing 5% (w/v) non-fat dry milk. Membranes were then washed five times with

TBST for 5 min before primary antibody incubation. The following antibodies were used for immunoblotting: polyhistidine (6x-His) tag-specific polyclonal antibody (Abcam, cat # ab9108; diluted 1:10,000), *E. coli* GroEL-specific antibody (Sigma-Aldrich, cat # G6532; diluted 1:10,000), and donkey anti-rabbit IgG conjugated to horseradish peroxidase (HRP) (Abcam, cat # ab7083; diluted 1:10,000). After probing with primary and second antibodies, the membranes were washed five times with TBST for 10 min. Chemiluminescence was developed using ECL substrate (Bio-Rad), and signals were imaged using a ChemiDoc MP Imaging System (Bio-Rad). Glycosylation efficiency was determined by performing densitometry analysis of protein bands in anti-His immunoblots using ImageJ software ⁴⁹ as described at <https://imagej.net>. Briefly, bands corresponding to g0 in each lane were grouped as a row or a horizontal “lane” and quantified using the gel analysis function in ImageJ. The bands corresponding to g1 were analyzed identically. The resulting intensity data for g0 and g1 was used to calculate percent glycosylated expressed according to the following ratio: $g1/[g0+g1]$. Efficiency data was calculated from immunoblots corresponding to three biological replicates, with all data reported as the mean $\pm$ SD. Statistical significance was determined by paired Student’s *t* tests (**p* < 0.05, ***p* < 0.01; ****p* < 0.001; *****p* < 0.0001) using Prism 10 for MacOS version 10.3.0.

***In vitro* glycosylation (IVG).** For *in vitro* glycosylation of scFv13-R4^{DQ^{NAT}}, 50 μ L of *in vitro* glycosylation buffer (50 mM HEPES (pH 7.4), 25 mM MnCl₂, and 0.1% (w/v) DDM) was supplemented with 1 μ g of scFv13-R4^{DQ^{NAT}} acceptor protein, 1–20 μ g of purified detergent-solubilized wt CjPglB or detergent-free water-soluble CjPglB, and 5–10 μ L of solvent-extracted CjLLOs (~10–20% (v/v) of the total reaction volume) and incubated at 30 °C for 16–20 h. Enzyme amounts were normalized based on molar concentrations rather than total protein mass to ensure accurate comparison of catalytic efficiencies. Organic solvent extraction of CjLLOs bearing the *C. jejuni* heptasaccharide glycan from the membrane of *E. coli* cells was performed as follows. *E. coli* CLM24 cells carrying plasmid pMW07-pgl Δ B were cultured overnight in LB medium supplemented with 20 μ g/mL chloramphenicol at 37 °C. Overnight cultures were subcultured into 1 L of TB medium with chloramphenicol and grown at 37 °C until the OD₆₀₀ reached $\approx$ 0.8. Protein expression was induced with 0.2% (v/v) L-arabinose, and cultures were incubated at 30 °C overnight. The next day, cells were harvested by centrifugation at 5,000 $\times$ g for 15 min at 4 °C, resuspended in methanol, and transferred

onto glass plates for drying overnight in a fume hood. The dried cell pellets were mechanically pulverized using a glass rod and subjected to lipid extraction with a 2:1 (v/v) chloroform:methanol mixture. Samples were sonicated in a water bath for 10 min and centrifuged at 3,000 ×g for 10 min at 4 °C. The supernatant was decanted, and the pellet was washed with water, followed by additional sonication and centrifugation. The resulting pellet was then extracted with a chloroform:methanol:water (10:10:3, v/v/v) mixture, sonicated again, and centrifuged at 3,000 ×g for 10 min at 4 °C. The upper aqueous phase was discarded, and the lower organic layer was collected into a fresh conical tube. After a final centrifugation at 3,000 ×g for 10 min at 4 °C, the organic phase was transferred to a clean glass plate and dried overnight in a fume hood. Dried CjLLOs were resuspended in a detergent-based buffer (50 mM HEPES, 0.1% (w/v) DDM, pH 7.4) and stored at −20 °C until further use in glycosylation assays.

Determination of enzyme kinetics. Enzyme kinetics were determined by performing IVG with TAMRA-labeled peptides as acceptor substrates. For turnover rate experiments, each reaction was prepared in a total volume of 100 μL *in vitro* glycosylation buffer supplemented with 1–2 μM of purified detergent-solubilized wt CjPgIB or detergent-free water-soluble CjPgIB, 5–10 μL (~10–20% (v/v)) of solvent-extracted CjLLOs, 0.5 μM of 5-TAMRA-GSDQNATF-NH₂ peptide (GenScript), and ddH₂O as needed. Reactions were incubated in a water bath at 30 °C with samples collected at different time points and stopped by boiling at 90 °C for 5 min. For Michaelis–Menten kinetics, reactions were performed in a total volume of 10 μL containing: 1 μL of *in vitro* glycosylation buffer, 0.5 μM of purified detergent-solubilized wt CjPgIB or detergent-free water-soluble CjPgIB, 5–10 μL (~10–20% (v/v)) of solvent-extracted CjLLOs, varying concentrations of 5-TAMRA-GSDQNATF-NH₂ (ranging from 0.1–2.0 μM), and ddH₂O as needed. The reactions were incubated for 90 min at 30 °C and stopped by boiling the sample at 90 °C for 5 min.

All reaction products were subjected to in-gel fluorescence analysis by diluting samples in Novex Tricine SDS Running Buffer (1x), mixing with dye that was produced in-house, and boiling at 80 °C for 2 min. The dye consisted of 200 mM Tris-Cl (pH 6.8), 8% (w/v) SDS (electrophoresis grade), and 40% (v/v) glycerol. Prior to gel loading, samples were diluted in Novex Tricine SDS Running Buffer (1x) so that concentration of total peptide in each lane was identical. A total of 5 μL of each sample was loaded onto Novex 16% Tricine Mini Protein Gels (1.0 mm thickness). The Spectra™

Multicolor Low Range Protein Ladder was used as the molecular weight marker. The gel was run at 70 V for 1.5 h at 4 °C to separate glycosylated peptides (g1) from non-glycosylated peptides (g0), and bands were subsequently visualized by fluorescence scanning of gels (ex: 488 nm; em: 526 nm) using a ChemiDoc MP Imaging System (Bio-Rad). DyLight 550 was used to visualize the fluorescently labeled peptides, while the Spectra ladder was visualized using Cy5.5. Using Prism 10 for MacOS (version 10.3.0), reaction velocities for each OST were determined by nonlinear regression analysis according to the Michaelis–Menten model while initial turnover rates were determined by linear regression analysis of the linear portion of the time-course reactions.

Acknowledgements. This work was supported by the Defense Advanced Research Projects Agency (DARPA contract W911NF-23-2-0039 to M.C.J. and M.P.D.), the Defense Threat Reduction Agency (grants HDTRA1-15-10052 and HDTRA1-20-10004 to M.P.D. and M.C.J.), the National Science Foundation (grants CBET-1605242 to M.P.D., CBET-1936823 and MCB-1413563 to M.P.D. and M.C.J.), and the National Institutes of Health (grants R01GM127578 to M.P.D.). L.M. is funded by HHMI Helen Hay Whitney postdoctoral fellowship. K.K. was supported by an American Heart Association Postdoctoral Fellowship. T.C.D. was supported by a Hartwell Foundation Postdoctoral Fellowship. E.J.B was supported by an NIH/NIGMS Chemical Biology Interface Training Grant (T32GM138826) and an NSF Graduate Research Fellowship (DGE-2139899). S.W.H. was supported by a training grant from the National Institutes of Health NIBIB (T32EB023860) and a Cornell Founder's Dissertation Fellowship. M.C. was supported by the National Science Foundation (DBI-2244288) through the Rosetta Commons REU. TJ was supported by an EMBO postdoctoral fellowship (ALTF 336-2021) and the Danish National Research Foundation (DNRF107).

Author Contributions. **Y.H.K.:** conceptualization, methodology, validation, formal analysis, investigation, writing – original draft, writing – review & editing, visualization. **L.M.:** conceptualization, methodology, validation, formal analysis, investigation, writing – original draft, writing – review & editing, visualization. **K.K.:** methodology, validation, formal analysis, investigation related to cloning, expression and purification of SIMPLEX constructs. **D.E.K.:** methodology and

investigation related to WRAP designs, writing – original draft, writing – review & editing. **T.C.D.:** methodology and investigation related to mass spectrometry analysis. **E.J.B.:** methodology and investigation related to mass spectrometry analysis. **C.K.B.:** methodology and investigation related to mass spectrometry analysis. **B.S.:** methodology and investigation related to enzyme kinetics. **S.W.H.:** methodology and investigation related to plasmid construction. **K.A.M.:** methodology and investigation related to conjugate vaccine expression and glycosylation. **A.T.:** methodology and investigation related to conjugate vaccine expression and glycosylation. **M.C.:** methodology and investigation related to WRAP designs. **D.M.:** conceptualization, methodology and investigation related to SIMPLEX designs. **T.J.:** conceptualization, methodology and investigation related to SIMPLEX designs. **H.C.:** supervision. **M.C.J.:** conceptualization, supervision, funding acquisition. **D.B.:** conceptualization, supervision, funding acquisition. **M.P.D.:** conceptualization, methodology, formal analysis, writing – original draft, writing – review & editing, visualization, supervision, project administration, funding acquisition.

Competing Interests Statement. M.P.D. and M.C.J. have financial interests in Gauntlet, Inc. and Resilience, Inc. M.P.D. also has financial interests in Cloaked Bio, Inc., Glycobia Inc., June Bio APS, UbiquiTx Inc., and Versatope Therapeutics Inc. M.P.D.’s and M.C.J.’s interests are reviewed and managed by Cornell University and Stanford University, respectively, in accordance with their conflict-of-interest policies. M.P.D. and M.C.J. declare no additional competing interests. Y.H.K. and M.P.D. are listed as co-inventors on a patent application filed by Cornell University related to the SIMPLEX method described in this paper. L.M., D.E.K. and D.B. are listed as co-inventors on a patent application filed by the University of Washington related to the WRAP method described in this paper. All other authors declare no competing interests.

References

1. Lairson, L.L., Henrissat, B., Davies, G.J. & Withers, S.G. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521-555 (2008).
2. Drula, E. et al. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* **50**, D571-D577 (2022).
3. Rini, J.M., Moremen, K.W., Davis, B.G. & Esko, J.D. in *Essentials of Glycobiology*, Edn. 4th. (eds. A. Varki et al.) 67-78 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY); 2022).
4. Moremen, K.W. & Haltiwanger, R.S. Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nat Chem Biol* **15**, 853-864 (2019).
5. Breton, C., Snajdrova, L., Jeanneau, C., Koca, J. & Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **16**, 29R-37R (2006).
6. Clausen, H., Wandall, H.H., DeLisa, M.P., Stanley, P. & Schnaar, R.L. in *Essentials of Glycobiology*, Edn. 4th. (eds. A. Varki et al.) 753-770 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY); 2022).
7. Jaroentomechai, T. et al. Cell-free synthetic glycobiology: designing and engineering glycomolecules outside of living cells. *Front Chem* **8**, 645 (2020).
8. Kightlinger, W., Warfel, K.F., DeLisa, M.P. & Jewett, M.C. Synthetic glycobiology: parts, systems, and applications. *ACS Synth Biol* **9**, 1534-1562 (2020).
9. Na, L., Li, R. & Chen, X. Recent progress in synthesis of carbohydrates with sugar nucleotide-dependent glycosyltransferases. *Curr Opin Chem Biol* **61**, 81-95 (2021).
10. Narimatsu, Y. et al. Genetic glycoengineering in mammalian cells. *J Biol Chem* **296**, 100448 (2021).
11. Natarajan, A., Jaroentomechai, T., Li, M., Glasscock, C.J. & DeLisa, M.P. Metabolic engineering of glycoprotein biosynthesis in bacteria. *Emerg Top Life Sci* **2**, 419-432 (2018).
12. Williams, A., Linhardt, R.J. & Koffas, M.A.G. Metabolic engineering of capsular polysaccharides. *Emerg Top Life Sci* **2**, 337-348 (2018).
13. Pandey, R.P., Parajuli, P. & Sohng, J.K. Metabolic engineering of glycosylated polyketide biosynthesis. *Emerg Top Life Sci* **2**, 389-403 (2018).
14. Schmid, J., Heider, D., Wendel, N.J., Sperl, N. & Sieber, V. Bacterial glycosyltransferases: challenges and opportunities of a highly diverse enzyme class toward tailoring natural products. *Front Microbiol* **7**, 182 (2016).
15. Moremen, K.W. et al. Expression system for structural and functional studies of human glycosylation enzymes. *Nat Chem Biol* **14**, 156-162 (2018).
16. Jaroentomechai, T. et al. A universal glycoenzyme biosynthesis pipeline that enables efficient cell-free remodeling of glycans. *Nat Commun* **13**, 6325 (2022).
17. Shimma, Y., Saito, F., Oosawa, F. & Jigami, Y. Construction of a library of human glycosyltransferases immobilized in the cell wall of *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **72**, 7003-7012 (2006).
18. Blount, Z.D. The unexhausted potential of *E. coli*. *Elife* **4** (2015).
19. Mizrahi, D. et al. Making water-soluble integral membrane proteins in vivo using an amphipathic protein fusion strategy. *Nat Commun* **6**, 6826 (2015).

20. Liu, J. & Mushegian, A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci* **12**, 1418-1431 (2003).
21. Gloster, T.M. Advances in understanding glycosyltransferases from a structural perspective. *Curr Opin Struct Biol* **28**, 131-141 (2014).
22. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K.P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350-355 (2011).
23. Shrima, S. & Gilmore, R. Oligosaccharyltransferase structures provide novel insight into the mechanism of asparagine-linked glycosylation in prokaryotic and eukaryotic cells. *Glycobiology* **29**, 288-297 (2019).
24. Kohda, D. Structural basis of protein Asn-glycosylation by oligosaccharyltransferases. *Adv Exp Med Biol* **1104**, 171-199 (2018).
25. Mihaljevic, L. et al. Solubilization of membrane proteins using designed protein WRAPS. *bioRxiv* (2025).
26. Jaroentomeechai, T. et al. Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat Commun* **9**, 2686 (2018).
27. Malakhov, M.P. et al. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics* **5**, 75-86 (2004).
28. Einhauser, A. & Jungbauer, A. The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. *J Biochem Biophys Methods* **49**, 455-465 (2001).
29. Borhani, D.W., Rogers, D.P., Engler, J.A. & Brouillette, C.G. Crystal structure of truncated human apolipoprotein A-I suggests a lipid-bound conformation. *Proc Natl Acad Sci U S A* **94**, 12291-12296 (1997).
30. Panagiotopoulos, S.E., Horace, E.M., Maiorano, J.N. & Davidson, W.S. Apolipoprotein A-I adopts a belt-like orientation in reconstituted high density lipoproteins. *J Biol Chem* **276**, 42965-42970 (2001).
31. Watson, J.L. et al. De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089-1100 (2023).
32. Dauparas, J. et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49-56 (2022).
33. Gerber, S. et al. Mechanism of bacterial oligosaccharyltransferase: in vitro quantification of sequon binding and catalysis. *J Biol Chem* **288**, 8849-8861 (2013).
34. Sotomayor, B. et al. Discovery of a single-subunit oligosaccharyltransferase that enables glycosylation of full-length IgG antibodies in bacteria. *Nat Commun* **16**, 6152 (2025).
35. Liu, F. et al. Rationally designed short polyisoprenol-linked PglB substrates for engineered polypeptide and protein N-glycosylation. *J Am Chem Soc* **136**, 566-569 (2014).
36. Ollis, A.A. et al. Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. *Sci Rep* **5**, 15237 (2015).
37. Kowarik, M. et al. N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* **314**, 1148-1150 (2006).
38. Ollis, A.A., Zhang, S., Fisher, A.C. & DeLisa, M.P. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat Chem Biol* **10**, 816-822 (2014).

39. Linton, D. et al. Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol Microbiol* **55**, 1695-1703 (2005).
40. Feldman, M.F. et al. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc Natl Acad Sci U S A* **102**, 3016-3021 (2005).
41. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345 (2009).
42. Hallgren, J. et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv* (2022).
43. Chaudhury, S., Lyskov, S. & Gray, J.J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689-691 (2010).
44. Vazquez Torres, S. et al. De novo design of high-affinity binders of bioactive helical peptides. *Nature* **626**, 435-442 (2024).
45. Bennett, N.R. et al. Improving de novo protein binder design with deep learning. *Nat Commun* **14**, 2625 (2023).
46. Leaver-Fay, A. et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* **487**, 545-574 (2011).
47. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
48. Lauer, T.M. et al. Developability index: a rapid in silico tool for the screening of antibody aggregation propensity. *J Pharm Sci* **101**, 102-115 (2012).
49. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).